Overcoming timescale and finite-size limitations to compute nucleation rates from small scale Well Tempered Metadynamics simulations

Matteo Salvalaglio,[1, a)] Pratyush Tiwary,[2] Giovanni Maria Maggioni,[3] Marco Mazzotti,[3] and Michele Parrinello[4, 5]

[1)] *Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom*

[2)] *Department of Chemistry, Columbia University, New York, New York 10027, United States of America*

[3)] *Institute of Process Engineering, ETH Zurich, CH-8092 Zurich, Switzerland*

[4)] *Department of Chemistry and Applied Biosciences, ETH Zurich, CH-8092 Zurich, Switzerland*

[5)] *Facoltà di informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland*

Condensation of a liquid droplet from a supersaturated vapour phase is initiated by a prototypical nucleation event. As such it is challenging to compute its rate from atomistic molecular dynamics simulations. In fact at realistic supersaturation conditions condensation occurs on time scales that far exceed what can be reached with conventional molecular dynamics methods. Another known problem in this context is the distortion of the free energy profile associated to nucleation due to the small, finite size of typical simulation boxes. In this work the problem of time scale is addressed with a recently developed enhanced sampling method while contextually correcting for finite size effects. We demonstrate our approach by studying the condensation of argon, and showing that characteristic nucleation times of the order of magnitude of hours can be reliably calculated, approaching realistic supersaturation conditions, thus bridging the gap between what standard molecular dynamics simulations can do and real physical systems.

I. INTRODUCTION

Nucleation is the event initiating first order phase transitions in which a small embryo of a thermodynamically stable phase appears within a parent metastable phase. The formation of gas bubbles in a liquid, liquid droplets in a vapour or crystal particles in solution are all examples of nucleation events playing a key roles in a variety of fields ranging from atmospheric physics to pharmaceutical manufacturing. The small length scale characterising nucleation events renders their direct experimental observation inherently challenging, while providing the ideal playground to apply and develop molecular modelling techniques. Despite extensive efforts in the investigation of nucleation phenomena with molecular simulations, the development of a systematic approach to the calculation of nucleation rates from first principles still remains a challenge, due to the very nature of the nucleation phenomena.

In the context of the condensation of a liquid phase from a supersaturated vapour, the nucleation rate J is defined to be the number of liquid droplets formed per unit time and volume.

The formation of a liquid droplet containing n molecules, in a system at constant volume (V) and temperature (T), is associated to a Helmholtz Free Energy change $\Delta F(n)$. The maximum of this quantity, ΔF^*, corresponding to a critical number of molecules n_d^*, constitutes the energy barrier that the system has to overcome in order to undergo the nucleation process. The very existence of this barrier determines the key features of nucleation, namely that it is an activated process and a paradigmatic example of a rare event.

The free energy barrier ΔF^* and the nucleation rate J depend on the thermodynamic driving force, that is typically expressed in terms of the supersaturation S, i.e. the ratio of the actual vapour pressure and the equilibrium vapour pressure. They depend also on temperature, T and on system-specific properties, namely the surface tension between the liquid and vapour phases, the molecular volumes of the two phases, v_ℓ and v_g, and the specific surface of the newly formed droplets, i.e. $6/d$ for a spherical droplet of diameter d.

In a system of volume V, at supersaturation S, and temperature T the characteristic time of nucleation τ can be expressed as:

$$\tau = \frac{1}{J(S,T)\,V} \qquad (1)$$

In Eq. 1 it can be readily seen that, at any given condition of temperature and supersaturation, τ, i.e. the average time necessary to observe a nucleation event, is inversely proportional to V. This relationship represents a constraint between two key factors determining the computational cost of MD simulations of nucleation events: the number of steps required to observe a nucleation event that needs to be of order of $\tau/\delta t$ where δt is the integration timestep used in MD, and the number of degrees of freedom which is instead proportional to the number of atoms, and thus to the volume V when comparing systems at the same density $\rho = N/V$. Such a constraint limits the range of conditions that could be directly investigated by small-scale MD simulations to regimes where $J \geqslant 10^{22} - 10^{25}$ cm^{-3} s^{-1}. Up to now,

---

[a)]Electronic mail: m.salvalaglio@ucl.ac.uk

overcoming such limitations has only been possible by using large scale simulations involving millions of atoms and requiring massive computational resources[1].

Nucleation is the prototypical example of a rare event, inherently stochastic in nature. When explicitly considering the stochastic character of nucleation τ represents the expected time for the nucleation of the first liquid droplet from a supersaturated vapour. The *law of rare events* suggests that the first nucleation event can be interpreted as a Poisson process[2], where the survival probability P_0, i.e. the probability that at a given time t there are no droplets in volume V , is:

$$P_0(t) = \exp\left(-\frac{t}{\tau}\right) \qquad (2)$$

Eq. 2 highlights how in order to reliably estimate τ, the stochastic nature of the nucleation process needs to be explicitly considered, while the reconstruction of the distribution of transition times is pivotal to reliably estimating J. In addition to the timescale issues, molecular simulations of nucleation processes suffer from intrinsic finite-size effects that cause a systematic distortion of $\Delta F(n)$ and impact rate calculations[3–8].

In this work we propose a systematic approach for the calculation of nucleation rates from small-scale unseeded molecular simulations. The proposed approach is based on recent developments of Well Tempered Metadynamics (WTmetaD)[9–11] that enable for the calculation of transition times distributions from biased simulations. Moreover we develop a systematic correction for the effect of finite-size on rate calculations.

Our method is tested on the paradigmatic case of the nucleation of a liquid argon droplet from a supersaturated argon vapour. The choice of this system has a two-fold aim: the first is to analyse a simple and yet very significant system that allows to draw conclusions general enough to be extended to a wider class of nucleation phenomena. The second is to point out that even in such a simple system, at realistic values of supersaturation the nucleation timescale rapidly grows out of reach of standard MD. Moreover choosing argon as a test case allows to benchmark our results against the existing literature based on small-scale unbiased MD simulations[12].

The paper is structured as follows; at first the details of the application of WTmetaD to the calculation of nucleation rates are reported, then an analysis of finite-size effects is carried out. WTmetaD and the finite size correction are then combined to outline a systematic strategy for the calculation of nucleation rates. Finally results are reported and commented. Unless otherwise noted, the subscript N will be used to refer to relevant quantities in finite-size systems. Such subscript will be dropped whenever referring to their counterpart in macroscopic systems.

II. NUCLEATION RATES AND LONG TIMESCALES

A. From metadynamics to dynamics

In this work the acceleration effect associated to WTmetaD has been exploited in order to substantially reduce the simulation time required to observe a nucleation event while simultaneously maintaining the system size small, hence significantly diminishing the overall computational cost. WTmetaD is conventionally used to compute free energy surfaces in a variety of contexts[10,11]. Recently it has been shown that, taking inspiration from conformational flooding[13] and hyperdynamics[14], transition times associated to activated events can be efficiently computed from WTmetaD simulations[15]. In WTmetaD the simulated system evolves in a transformed time coordinate, t_{WT}, due to the application of the history-dependent bias potential $V_B(\xi, t)$ constructed as a function for the collective variable ξ[10,15]. As discussed in detail in Ref. 15, rate calculations via WTmetaD do not require a converged estimate of free energy profiles, being instead based on the systematic evaluation of the so-called *acceleration factor*, which represents the ratio between the physical time and the meta dynamics time.

In the context of a nucleation problem, the specific transition time associated to a nucleation event t_n can be computed from the corresponding WTmetaD simulation time $t_{n,WT}$ as:

$$t_n = t_{n,WT} \langle \exp\left(\beta V_B(\xi, t)\right) \rangle_{WT} \qquad (3)$$

where the term $\langle \exp\left(\beta V_B(\xi, t)\right) \rangle_{WT}$ is the *acceleration factor* called α in the following. Note that $\beta = 1/k_B T$. Crucial to this procedure is the hypothesis of negligible bias deposition at the transition state. To comply with such hypothesis, the bias potential is constructed through the *infrequent* deposition of potential Gaussians, in a properly defined space of collective variables. The fulfillment of such condition can be checked *a posteriori* using the approach detailed in Ref. Salvalaglio2014assessing. When this is the case, the whole transition time distribution can be recovered from a set of WTmetaD simulations. This approach has been applied to several problems, thus allowing the computation of rates of activated processes such as DNA unfolding[16], and protein-ligand unbinding[17–19].

B. A collective variable to describe liquid argon nucleation

In WTmetaD the bias potential $V_B(\xi, t)$ is constructed as a function of a collective variable ξ[10,11,20]. In this work we choose as collective variable n, the total number of liquid argon atoms in the system, corresponding to the typical reaction coordinate used in classical nucleation theory (CNT). In order to be used in the framework of WTmetaD, n has been expressed as a continuous and differentiable function of the atomic coordinates[9,10,21]. To

this aim the ten Wolde-Frenkel definition[12,22] has been applied, in which atoms are considered liquid when they possess a coordination number larger than a threshold value c_ℓ, that is chosen to be 5. The coordination number of each molecule in the system is defined in a continuous and differentiable form through the expression: $c_i = \sum_{j \neq i} f(r_{ij})$, where r_{ij} is the Cartesian distance between atoms i and j and $f(r_{ij})$ is the switching function:

$$f(r_{ij}) = \frac{1 - (r_{ij}/r_c)^6}{1 - (r_{ij}/r_c)^{12}} \tag{4}$$

The number of molecules possessing $c_i \geq c_\ell$ is thus calculated using the same functional form as in Eq. 4:

$$n = \sum_{i=1}^{N} \frac{1 - (c_\ell/c_i)^6}{1 - (c_\ell/c_i)^{12}} \tag{5}$$

C. Detecting nucleation events in WTmetaD simulations

The biased transition time $t_{n,WT}$ is the simulation time associated with the occurrence of a nucleation event in a WTmetaD simulation. Hence to compute $t_{n,WT}$, it is necessary to reliably detect nucleation events. In our case we apply the approach described in Ref. 12 which is based on the fact that a clear timescale separation exists between the residence time in the supersaturated vapour state and the time necessary for a supercritical nucleus to grow in size. The latter phenomenon is orders of magnitude faster than the former, rendering nucleation a rare but fast event. As done in Ref. 12, the nucleation time $t_{n,WT}$ can be directly calculated from the time evolution of $n(t)$, as the simulation time needed to overcome a threshold size of the emerging liquid droplet $\overline{n}$. This approach remains valid as long as the threshold size chosen to define the transition criterion is larger than the critical nucleus. In such case the transition time can be safely considered independent of the specific value of $\overline{n}$[12].

D. Expected nucleation time and nucleation rates

As briefly mentioned in the introduction, due to the activated nature of nucleation, its transition time probability distribution is described by the so-called *law of rare events*, and is thus expected to be exponential. The nucleation process, particularly the formation of the first nucleus, i.e. the event that matters at the scale of the MD simulation box, can in fact be modelled by a time-homogeneous Poisson process characterised by a survival probability $P_0(t) = \exp(-t/\tau_N)$. Due to its inherent stochasticity, an appropriate sampling of the nucleation times distribution is required to evaluate the expected characteristic nucleation time τ_N[12,23,24]. In order to compute the expected nucleation time τ_N in a finite-size system, we perform a large number of independent WTmetaD NVT simulations and extract from each of them

a nucleation time t_N. After this, the survival probability distribution constructed from the t_N values is analysed and its statistical compatibility with a Poisson process quantified. This allows to check whether the conditions under which Eq. 3 is valid are satisfied[25]. The characteristic nucleation time fitted from the survival probability distribution τ_N is thus used to compute the nucleation rate in the finite size system as $J_N = (\tau_N V)^{-1}$, where V is the system's volume[12,23,24].

E. WTmetaD simulation details

Transition times were computed from NVT simulations of systems consisting of 512 argon atoms at a constant temperature of 80.7 K. A Lennard-Jones potential with $\epsilon = 0.99797$ kJ/mol and $\sigma = 0.3405$ nm has been adopted to describe the interactions between argon atoms[12]. The potential has been truncated, but not shifted, with a cut-off length of $6.75\,\sigma$. The time step for the integration of the equations of motion has been set to 5 fs[12]. These specific conditions were chosen in order to compare our results with those reported in Ref. 12. The equilibrium vapour pressure (p_e) of argon under these conditions is equal to 0.43 bar[12,26]. To carry out rate calculations in a wide supersaturation range we have followed the approach proposed in Ref. 12, carrying out a series of NVT simulations in cubic boxes of increasing volume, corresponding to supersaturation values ranging from 11.4 to 5.4.

TABLE I. WTmetaD simulation setup summary.

Label	S	l	p	n_{sim}	σ	ω_0	Δt	γ	$V_B^0(nl)$
		[nm]	[bar]			[kJ/mol]	[ps]		[Y/N]
S_1	11.4	10.5	4.86	100	0.5/1.0	0.01	25	5	N
S_2	8.68	11.5	3.70	100	0.5	0.01	25	5	N
S_3	6.76	12.5	2.88	50	1.0	0.02	25	5	N
S_4	6.01	13.0	2.56	50	0.5	0.01	25	5	Y
S_5	5.36	13.5	2.28	50	0.5	0.01	25	5	Y

In Tab. I setup parameters of the five sets of simulations, namely the supersaturation level S, dimension of the simulation box edge l, initial pressure p, and number of independent simulations per supersaturation level n_{sim} have been reported. In Tab. I we also report the WTmetaD setup parameters, namely width of the deposited Gaussians δ, their initial height ω_0, the deposition stride Δt and the γ factor. For a detailed description of the metadynamics algorithm and parameters the interested reader is invited to check the references[9–11]. In the last column we indicate whether an initial bias potential $V_B^0(n)$ was applied (Y) or not (N) (see the results section for a description of simulations with and without $V_B^0(n)$). The highest supersaturation at which we

have performed simulations is $S = 11.4$, corresponding to the lowest supersaturation at which the standard simulations of Ref.[12] were performed. This allowed us to check that our simulation setup was correctly reproducing nucleation rates both in biased and unbiased simulations. For each supersaturation the survival probability distribution has been constructed by performing 50 to 100 independent nucleation simulations. The bias was updated every 5000 integration steps, with a bias-factor of 5, and a Gaussian height of 1×10^{-2} or 2×10^{-2} kJ/mol. The values of the collective variable n and of the total bias $V_B(n,t)$ have been collected every 100 steps. Temperature has been controlled using the Bussi-Donadio-Parrinello thermostat[27], with a time constant of 0.1 ps. WTmetaD simulations were performed with Gromacs 4.6.3[28] equipped with PLUMED 2.0[29].

III. NUCLEATION RATES AND FINITE SIZE EFFECTS

The nucleation rate of droplets from a vapour can be explicitly derived within classical nucleation theory (CNT) as:

$$J = A \left(\frac{p}{p_e} \right) \exp \left(-\beta \Delta F^* \right) \tag{6}$$

where A is a pre-exponential factor, p is the pressure in the vapour phase, p_e the equilibrium vapour pressure, and ΔF^* the free energy barrier to nucleation.
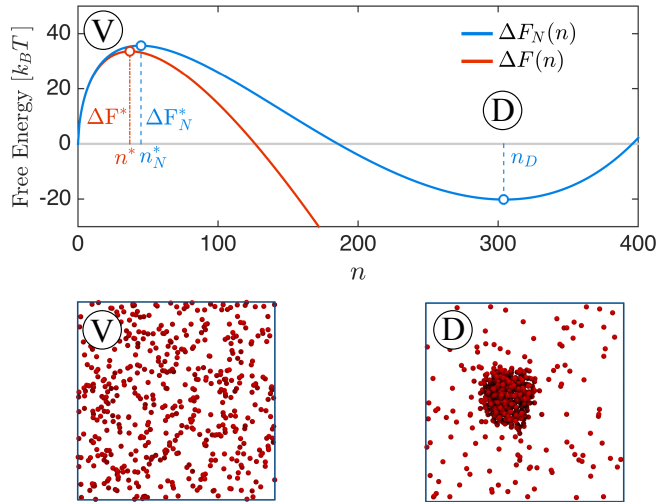

FIG. 1. Nucleation of a liquid argon droplet from supersaturated vapour. *top* Free energy profiles predicted by CNT in an infinitely large system at constant supersaturation ($\Delta F(n)$) and for a finite-size, confined system ($\Delta F_N(n)$). Both free energy profiles refer to a system of 512 argon atoms, in a volume of 2197 nm^3, at supersaturation $S=6$, with a surface energy $\sigma a=9$ $k_B T$. *bottom* Representation of the system in its vapour V and droplet D configurations.

Equation 6 can be viewed as the product of two distinct contributions: an energetic part, corresponding to the exponential term, and a kinetic one, related to the molecular collisions and given by the pre-exponential term. In order to derive a systematic correction to nucleation rates computed from small-scale molecular simulations, in the following we shall assess the impact of the finite size of the simulation box on both terms.

A. Free energy of nucleation in a confined system

We define confinement as the impossibility of exchanging atoms between the system and the surrounding environment. Under this definition a NVT simulation box, represents a prototypical confined system, as its total number of atoms N is by definition constant. As highlighted in several reference works[3–8], the free energy change associated with a nucleation process is affected by confinement. In Fig. 1, the comparison between nucleation free energies in finite-size (blue) and macroscopic (red) conditions is illustrated together with the representation of typical configurations of Argon atoms in the vapour (V) and liquid droplet (D) states. Hereafter we shall summarise how such a distortion in the free energy profile affects nucleation rates[3]. We treat the Argon vapour as an ideal gas hence the pressure of the vapour before droplet formation (initial state of the system) is:

$$p_0(N, V, T) = \frac{N k_B T}{V} \tag{7}$$

A liquid spherical embryo of n molecules has a volume V_d, and a surface A_d, that can be expressed as a function of n as:

$$V_d = n v_\ell \tag{8}$$

$$A_d = (an)^{2/3} \tag{9}$$

where v_ℓ is the molecular volume in the liquid phase, and $a = 6\pi^{1/2} v_\ell$. After formation of such an embryo (the final state of the system) the vapour pressure attains the following value:

$$p(N, n, V, T) = \frac{(N - n) k_B T}{V - n v_\ell} \tag{10}$$

The Helmholtz free energy change, ΔF_N, for the transition from the initial to the final state, i.e. for the formation at constant temperature T and volume V of a n-molecule droplet from a vapour consisting initially of N molecules, is given by[4]:

$$\Delta F_N (N, n, V, T) = -n\beta^{-1} \ln \left(\frac{p}{p_e} \right) + \gamma (an)^{2/3} +$$
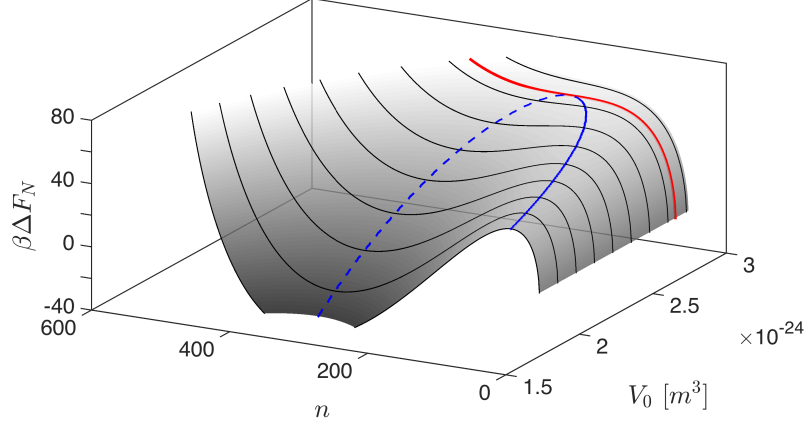$$+ N\beta^{-1} \ln \left(\frac{p_0}{p_e} \right) + n \left(\beta^{-1} - v_\ell p_e \right) \tag{11}$$

FIG. 2. Free energy of nucleation in a finite-size system at constant N=512, T=80.7 K, as a function of the system volume V. The solid blue line represents the locus of the maxima of ΔF_N, corresponding to the critial nuclei. The dashed blue line represents the locus of the local minima in of ΔF_N, representing a stable argon droplet in a finite-size induced equilibrium with the argon vapour. In solid red the $\Delta F(n)$ is highlighted, corresponding to the threshold value of V above which $\Delta F(n)$ becomes a monotonically increasing function even if $S_0 > 1$.

where volume, surface, and pressure effects are accounted for. A typical $\Delta F_N(n)$ profile is reported in Fig. 1 (a) (blue). Both a local maximum and a local minimum can be identified along $\Delta F_N(n)$. The local maximum ΔF_N^* represents the free energy barrier to nucleation in a finite size system. ΔF_N^* is associated with a critical nucleus size n_N^*, that represents the liquid embryo in unstable equilibrium with the surrounding vapour. Such ΔF_N^* value can be computed numerically. As extensively discussed in Ref. 3 for an argon vapour and in Refs. 7, 8 for the case of crystal nucleation from solution, there exists a minimum value of the initial supersaturation, $S_0 = p_0/p_e > 1$, below which the function $\Delta F_N(n)$ is monotonically increasing, hence no maximum is present. At N and T fixed such a condition defines an upper bound for the volume for which nucleation rates can be computed. In Fig. 2 $\Delta F_N(n)$ is plotted as a function of the system volume, highlighting the critical conditions associated with the transition to a monotonically increasing function. Contrary to the analysis so far, Classical Nucleation Theory (CNT) deals with infinitely large systems, where the formation of the liquid droplet has a negligible effect on the state of the surrounding phase. The corresponding Helmholtz free energy change, $\Delta F(n)$, for the formation of a n-molecule embryo at temperature T and supersaturation S_0 can be obtained by taking the limit of Eq. 11 with V and N approaching infinity, and their ratio remaining constant. Under these conditions $p = p_0$ and:

$$\Delta F(n) = -n k_B T \ln S_0 + \gamma (an)^{2/3} \qquad (12)$$

For an infinitely large system, in which Eq. 12 holds the free energy barrier ΔF^* can be computed analytically as[30]:

$$\Delta F^* = -\frac{4\beta^3 \gamma^3 a^2}{27 (\ln S_0)^2} \qquad (13)$$

It is worth noticing that ΔF_N^*, which is the nucleation free energy barrier in a confined system at the same conditions of T and S_0, is strictly larger than ΔF^*.

B. Macroscopic nucleation rates from finite size calculations

In order to compute a correction term associated to the confinement effect we follow the approach of Ref. 3, and define a factor ϕ as the ratio between the nucleation rate in macroscopic conditions J and in a finite sized confined system J_N. In analogy with Eq. 6, the nucleation rate in a confined system J_N can be written as[4]

$$J_N = A_N S \exp\left(-\beta \Delta F_N^*\right) \qquad (14)$$

Since the system supersaturation in a finite-size simulation before a supercritical nucleus forms is essentially the same as in a macroscopic system, $S = S_0$, ϕ reduces to[3]:

$$\phi = \frac{J}{J_N} = \frac{A}{A_N} \exp\left(\beta(\Delta F_N^* - \Delta F^*)\right) \qquad (15)$$

Eq. 15 provides a working principle to obtain nucleation rates in macroscopic systems from finite size NVT simulations as:

$$J = \phi J_N \qquad (16)$$

As reported in Ref. 3, due to the exponential dependence on the strictly positive quantity $\Delta F_N^* - \Delta F^*$, the
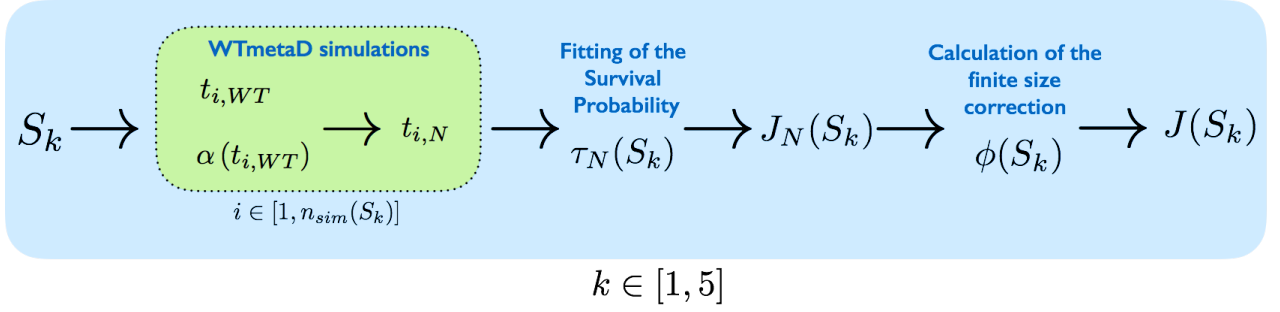
FIG. 3. Workflow summary for the calcuations of macroscopic nucleation rates from finite size NVT WTmetaD nucleation simulations.

dominating term in Eq. 15 is $\exp\left(\beta(\Delta F_N^* - \Delta F^*)\right)$, whereas the pre-exponential term A/A_N in Eq. 15 is expected to play a secondary role. In order to identify key contributions to A/A_N, its dependence on finite size is discussed in the following.

Within the framework of CNT, A is typically expressed as[12,30]

$$A = Z f^* \frac{p_e}{k_B T} \qquad (17)$$

where Z is the Zeldovich factor given by[31]:

$$Z = \sqrt{\frac{\left|\frac{d^2 \Delta F(n)}{dn^2}\big|_{n=n^*}\right|}{2\pi k_B T}} \qquad (18)$$

and f^* the rate of attachment of molecules to the critical cluster. Since nucleation of a droplet from its vapour is a process controlled by direct impingement[30], the attachment rate f^* is derived from the kinetic theory of gases as[30,32]:

$$f^* = c(n^*) \frac{p}{\sqrt{2\pi m k_B T}} \qquad (19)$$

where $c(n^*) = \sqrt[3]{(36\pi v_\ell^2)}(n^*)^{2/3}$ is the surface area of the critical cluster, v_ℓ is the volume per molecule in the liquid phase and p the pressure.

The attachment frequencies f^* and f_N^* differ due to two reasons. The first is that the critical nucleus size in finite size simulations n_N^* is strictly larger than the critical nucleus in the corresponding infinite case n^{*4}. The second reason is that the vapour pressure acting on the critical nucleus in finite size systems $p_N^* = (N - n_N^*)/((V - n_N^* v_\ell)k_B T)$ is always smaller than its corresponding value for a system at macroscopic conditions $p = N/(V k_B T)$.

The Zeldovich factors Z and Z_N are instead expected to differ due to the fact that the curvature of the free energy profile in the region around its maximum is affected by finite size, see Fig. 1 for example. The extent of the contribution of the $\exp\left(\beta(\Delta F_N^* - \Delta F^*)\right)$, f^*/f_N^*, and Z/Z_N terms to ϕ is discussed in the Results section.

The correction factor ϕ depends on quantities that can be directly calculated from $\Delta F(n)_N$ and $\Delta F(n)$ such

as: ΔF_N^*, ΔF^*, Z_N, Z, n_N^*, and n^*. Both $\Delta F(n)$ and $\Delta F(n)_N$ can be respectively computed from Eq. 12 and 11, once the surface tension γ is known.

The surface tension γ is obtained by fitting Eq. 14, in which the pre-exponential term A_N in considered supersaturation-independent, on the J_N values obtained as a function of supersaturation.

IV. WORKFLOW SUMMARY

In Fig. 3 the workflow for the calculation of nucleation rates from small-scale finite size NVT simulations has been summarised. The calculation procedure can be outlined as follows:

1. A set of WTmetaD simulations is carried out for multiple supersaturation levels. Supersaturation is imposed by defining the system volume while keeping constant the number of molecules N and the temperature T.

2. Applying the criterion for the identification of nucleation events proposed in Ref. 12, the WTmetaD transition time t_{WT} and the corresponding acceleration factor α are calculated from each WTmetaD simulation.

3. The physical transition time associated to each nucleation event is computed using Eq. 3.

4. The transition times obtained for each supersaturation value are used to fit the survival probability distribution, and compute the average nucleation time τ_N for each finite size system at volume V and supersaturation S.

5. Average nucleation times τ_N are converted to finite-size nucleation rates using $J_N = 1/(\tau_N V)$[12].

6. The finite-size nucleation rates are used to fit Eq. 14. The fitting parameter is the surface tension γ, which is used to compute $\Delta F*$, ΔF_N^*, n^*, n_N^*, Z, Z_N, and thus the correction factor ϕ.

7. Eq. 16 is used to compute the nucleation rate in macroscopic conditions J.

V. RESULTS

A. WTmetaD simulations

The time evolution of the number of liquid-like argon atoms n in a typical WTmetaD simulation is reported in Fig. 4, where the nucleation event can be clearly identified as the rapid transition from n values fluctuating close to zero, to n values fluctuating around a positive value n_D. The final state corresponds to a finite-sized droplet stabilised by finite size effects corresponding to the local minimum in free energy shown in Fig. 1 (blue curve).
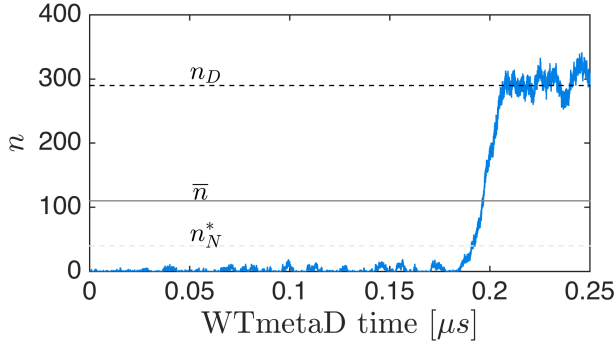
FIG. 4. Time series of the collective variable $n(t)$ obtained from a typical simulation. The value of typical $\overline{n}$, n_N^*, and n_D have been highlighted on the plot.

It can be seen that the lifetime of the supersaturated vapour state in the WTmetaD simulation is much larger than the transition time associated with the nucleation event driving the system into the stable state characterised by $n = n_D$. Such a difference becomes exponentially large when the WTmetaD time is rescaled to real time according to Eq. 3[15], neatly highlighting the timescale separation characteristic of the nucleation problem. During WTmetaD simulations a repulsive bias potential is adaptively constructed with an infrequent deposition of Gaussians[9,15]. In order to speed up the adaptive construction of the bias for the two slowest cases (S_4 and S_5 in Tab . I), in addition to the WTmetaD bias $V_B(n,t)$, we apply a static bias $V_B^0(n)$ constructed from a preliminary WTmetaD simulation. In Fig. 5 the total bias potential is reported for supersaturation levels S_2 and S_5, which are characterised by the absence and presence of an initial bias $V_B^0(n)$, respectively. In all cases in the region of the maximum of $\Delta_N F(n)$, the total bias applied $V_B^{tot}(n)$ decays to values smaller than $k_B T$, in agreement with the hypotheses of negligible bias deposition at the transition state invoked to carry out rate
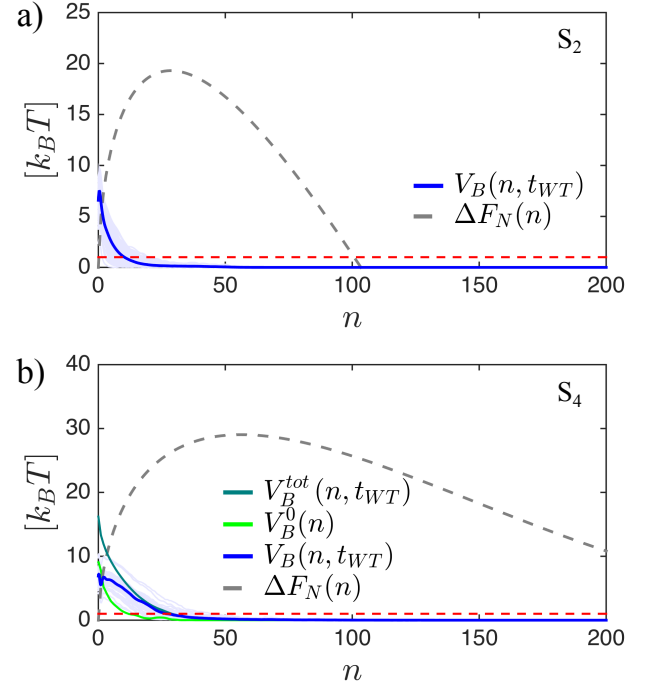
calculations from WTmetaD[15,25].

FIG. 5. a) Simulation set S_2: WTmetaD bias potential at transition time, $V_B(n,t_{WT})$. b) Simulation set S_4: initial static bias potential $V_0(n)$, WTmetaD bias potential at transition time $V_B(n,t_{WT})$, and total bias at transition time $V_B^{tot}(n,t_{WT}) = V_B^0(n) + V_B(n,t_{WT})$. For comparison nucleation free energy profiles $\Delta F_N(n)$ have been reported and the $k_B T$ level has been highlighted in red.

B. Survival probability distributions and average transition times in finite size conditions

As described in section II, sets of 50 to 100 simulations were carried out at five different supersaturation levels S_1-S_5 (see I). From each set of simulations an *empirical* survival probability distribution (ESP) has been constructed. The average transition time in finite-size conditions τ_N has been computed for each supersaturation level by a non-linear least square fitting of the ESP with the expression $P_0 = \exp(-t/\tau_N)$, hereafter referred to as the theoretical survival probability (TSP). Evaluating the statistical compatibility between the ESP and the TSP with the protocol described in Ref. 25, allowed ensuring that the crucial hypothesis of negligible bias deposition at the transition state has been satisfactorily fulfilled for all the simulation sets S_1-S_5. In Fig.6a both the ESP constructed from WTmetaD simulations and the fitted TSP are reported; in Fig.6b the supersaturation-dependent values of the finite-size transition time τ_N is reported as a function of supersaturation, together with the average acceleration factor α (in the inset). Fig. 6b
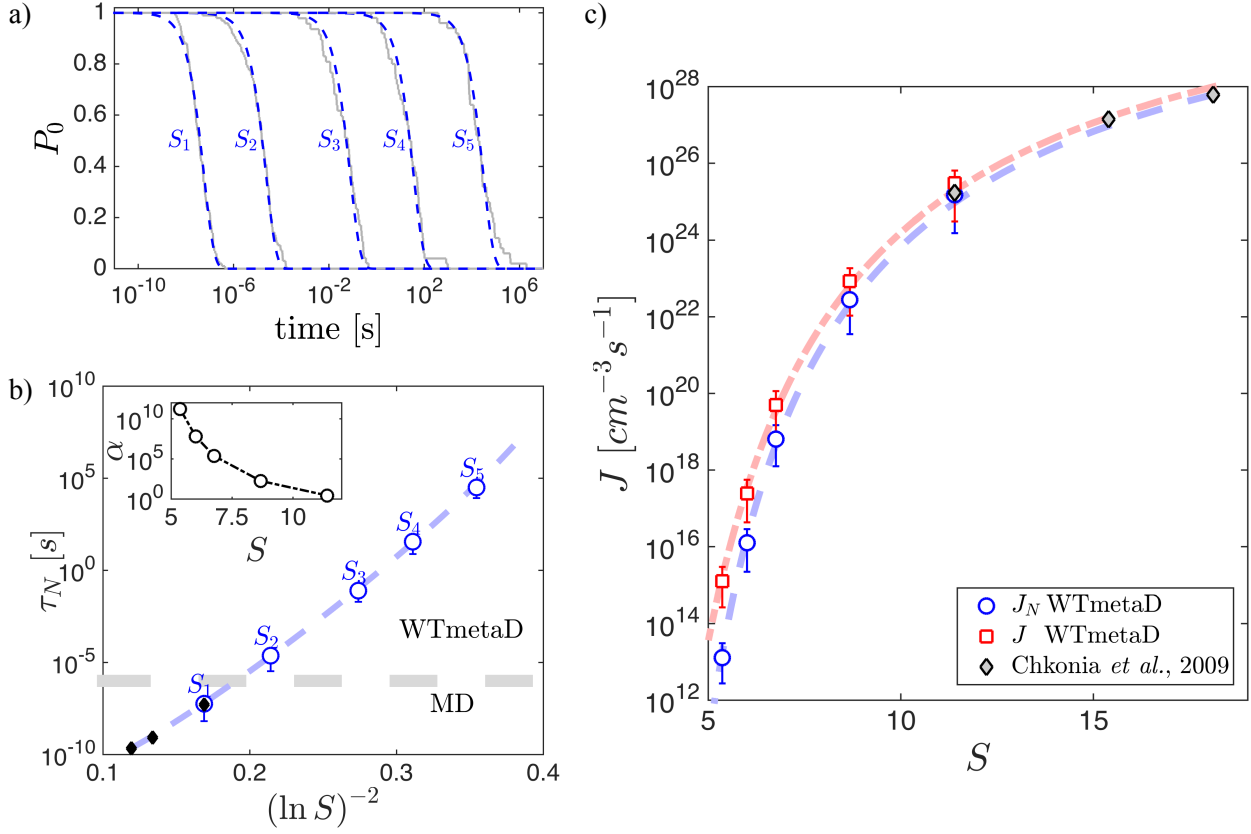
FIG. 6. *a)* Survival probability distributions obtained for different supersaturations (S_1=11.4, S_2=8.7, S_3=6.8, S_4=6.0, S_5=5.4). *b)* Average nucleation times in finite size systems τ_N. The average acceleration factor α as a function of supersaturation is displayed as an inset. *c)* Nucleation rates calculated from finite-sized WTmetaD simulations (J_N) and rescaled to the macroscopic limit J. The blue dashed line represents the result of the fitting of the J_N data with Eq. 14. Computed values of J, J_N, τ_N, α, convergence of the τ_N values and their associated errorbars are reported in the Supplementary Information (SI).

highlights how the application of WTmetaD allows to directly simulate nucleation events characterised by transition times of the order of 10^4 seconds, thus significantly expanding the range of transition times that could be reached with standard MD in a similar simulation setup.

C. Nucleation rates

As discussed in section II D nucleation rates in the confined regime J_N can be directly computed as $J_N = (\tau_N V)^{-1}$. In Fig. 6c values of J_N are reported as a function of the supersaturation S, clearly showing a significant extension of the accessible nucleation rates domain, which is increased up to ten orders of magnitude compared to the domain typically accessible by standard molecular dynamics[12,23,24,33,34].

The surface tension γ obtained from the fitting of the J_N values computed from WTmetaD corresponds to $\gamma = 16.9$ mN/m, a value that nicely extrapolates the data of Goujon *et al.*[35] for the same system in the temperature range between 85 K and 135 K as shown in the Supplementary Information (SI). The surface tension has

been considered independent from S, as typically done in CNT. We have found that this choice allows to well describe the J_N data directly computed from simulations while keeping at a minimum the number of fitting parameters. We have also verified that considering γ a linear function of supersaturation does not noticeably improve the description of the WTmetaD data.

In Fig. 7, it can be seen that the finite size correction is negligible at high supersaturation ($S \geq 11.4$), where the nucleation barriers ΔF^* and ΔF_N^* are almost indistinguishable. However, for lower supersaturation levels ϕ reaches values accounting for up to two orders of magnitude of difference between J and J_N. This finding highlights the importance of properly handling finite size effects when investigating transitions at low supersaturations. A breakdown of the contributions of the terms appearing in ϕ is also reported in Fig. 7. It can be seen that, as expected, the contribution of the term f^*/f_N^* negligible over the entire supersaturation domain. Despite the term Z/Z_N having a slightly heavier impact on ϕ, it can be seen that the finite-size correction is substantially captured by considering only the exponential term in Eq. 15.
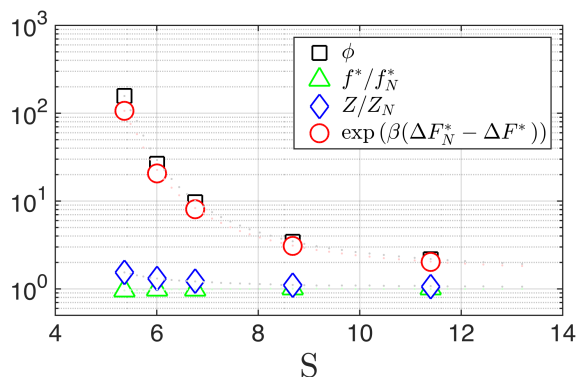
FIG. 7. Breakdown of the contributions to the finite size correction ϕ of the factors $\exp\left(\beta(\Delta F_N^* - \Delta F^*)\right)$, f/f_N, and Z/Z_N.

Nucleation rates in macroscopic conditions J are thus computed as:

$$J = \phi J_N \simeq J_N \exp\left(\beta\left(\Delta F_N^* - \Delta F^*\right)\right) \qquad (20)$$

and reported in Fig. 6c. Nucleation rates rescaled explicitly accounting also for the term Z/Z_N are reported in the SI.

VI. CONCLUSION

To conclude, in this work we have shown that WT-metaD can be applied to the direct calculation of nucleation rates, proving to be particularly useful to tackle the timescale limitations that plague small-scale nucleation simulations. In the case of argon condensation this implies being able to simulate nucleation in fairly small systems, while efficiently reaching timescales of the order of 1×10^4 s with ordinary computational resources. Moreover, we have highlighted that rate calculations from small scale simulations require a systematic assessment of finite-size effects. Being able to simultaneously address both timescale and finite-size limitations allows to significantly extend the range of nucleation conditions that can be directly investigated with computationally efficient, small-scale molecular simulations.

[1] J. Diemand, R. Angelil, K. K. Tanaka, and H. Tanaka, J. Chem. Phys. **139**, 074309 (2013).

[2] S. I. Resnick, *Adventures in stochastic processes* (Springer, 1992).

[3] J. Wedekind, D. Reguera, and R. Strey, J. Chem. Phys. **125**, 214505 (2006).

[4] D. Reguera, R. K. Bowles, Y. Djikaev, and H. Reiss, J. Chem. Phys. **118**, 340 (2003).

[5] J. W. W.P. Schmelzer and A. S. Abyzov, J. Non-Cryst. Solids **384**, 2 (2014).

[6] R. Grossier and S. Veesler, Cryst. Growth Des. **9**, 1917 (2009).

[7] M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello, Proc. Nat. Acad. Sci. **112**, E6 (2015).

[8] M. Salvalaglio, M. Mazzotti, and M. Parrinello, Faraday Discuss. **179**, 291 (2015).

[9] A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).

[10] A. Barducci, M. Bonomi, and M. Parrinello, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 826 (2011).

[11] O. Valsson, P. Tiwary, and M. Parrinello, Annu. Rev. Phys. Chem. **67**, null (2016), pMID: 26980304.

[12] G. Chkonia, J. Wolk, R. Strey, J. Wedekind, and D. Reguera, J. Chem. Phys. **130**, 064505 (2009).

[13] H. Grubmüller, Phys. Rev. E **52**, 2893 (1995).

[14] A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).

[15] P. Tiwary and M. Parrinello, Phys. Rev. Lett. **111**, 230602 (2013).

[16] F. Sicard, N. Destainville, and M. Manghi, J. Chem. Phys. **142**, 034903 (2015).

[17] P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, Proc. Nat. Acad. Sci. **112**, E386 (2015).

[18] P. Tiwary, J. Mondal, J. A. Morrone, and B. J. Berne, Proc. Nat. Acad. Sci. **112**, 12015 (2015).

[19] J. I. Stuckey, B. M. Dickson, N. Cheng, Y. Liu, J. L. Norris, S. H. Cholensky, W. Tempel, S. Qin, K. G. Huber, C. Sagum, *et al.*, Nat. Chem. Biol. **12**, 180 (2016).

[20] J. F. Dama, M. Parrinello, and G. A. Voth, Phys. Rev. Lett. **112**, 240602 (2014).

[21] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. USA **99**, 12562 (2002).

[22] P. R. ten Wolde and D. Frenkel, The Journal of Chemical Physics **109** (1998).

[23] J. Wedekind, J. Wolk, D. Reguera, and R. Strey, J. Chem. Phys. **127**, 154515 (2007).

[24] D. Yuhara, B. C. Barnes, D. Suh, B. C. Knott, G. T. Beckham, K. Yasuoka, D. Wu, and A. K. Sum, Faraday Discuss. **179**, 291 (2015).

[25] M. Salvalaglio, P. Tiwary, and M. Parrinello, J. Chem. Theory Comput. **10**, 1420 (2014).

[26] V. G. Baidakov, S. P. Protsenko, Z. R. Kozlova, and G. G. Chernykh, J. Chem. Phys. **126**, 214505 (2007).

[27] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

[28] E. Lindahl, B. Hess, and D. Van Der Spoel, J. Mol. Model. **7**, 306 (2001).

[29] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comp. Phys. Commun. **185**, 604 (2014).

[30] D. Kashchiev, *Nucleation, Basic Theory with Applications*, edited by Elsevier (Elsevier, 2000).

[31] Y. B. Zeldovich, Acta physicochim. URSS **18**, 1 (1943).

[32] J. Wedekind, J. Wölk, D. Reguera, and R. Strey, J. Chem. Phys. **127**, 154515 (2007).

[33] K. K. Tanaka, A. Kawano, and H. Tanaka, J. Chem. Phys. **140**, 114302 (2014).

[34] F. Zipoli, T. Laino, S. Stolz, E. Martin, C. Winkelmann, and A. Curioni, J. Chem. Phys. **139**, 094501 (2013).

[35] F. Goujon, P. Malfreyt, and D. J. Tildesley, J. Chem. Phys. **140** (2014), 10.1063/1.4885351.

Overcoming timescale and finite-size limitations to compute nucleation rates from small scale Well Tempered Metadynamics simulations: Supplementary Information.

Matteo Salvalaglio,[1, a)] Pratyush Tiwary,[2] Giovanni Maria Maggioni,[3] Marco Mazzotti,[3] and Michele Parrinello[4, 5]

[1)] *Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom*

[2)] *Department of Chemistry, Columbia University, New York, New York 10027, United States of America*

[3)] *Institute of Process Engineering, ETH Zurich, CH-8092 Zurich, Switzerland*

[4)] *Department of Chemistry and Applied Biosciences, ETH Zurich,CH-8092 Zurich, Switzerland*

[5)] *Facoltà di informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland*

I. TRANSITION TIMES: TABLES AND CONVERGENCE

Average transition times τ_N, rates in the confined system J_N, and corrected, macroscopic rates rates J are reported in Tab. S1. In order to measure the compatibility between the empirical survival probability distribution computed from WTmetaD simulations and the exponential distribution theoretically predicted by the law of rare events a Kolmogorov-Smirnov (KS) test has been carried out[1]. The p-value associated to the KS statistic is reported in Tab. S1. The p-value, in this context, represents the probability that the survival probability distribution obtained from WTmetaD obeys the theoretical exponential distribution. In all cases the p-value is well above the standard significance level of 0.05. In Ref. 1 this analysis approach is validated and discussed at length.

Label	τ_N [s]	J_N [cm^{-3} s^{-1}]	J [cm^{-3} s^{-1}]	*p-value*	α
S_1	$5.75 \pm 0.65 \times 10^{-8}$	$1.5 \pm 0.34 \times 10^{25}$	$3.04 \pm 0.70 \times 10^{25}$	0.84	2.8
S_2	$2.33 \pm 0.33 \times 10^{-5}$	$2.8 \pm 0.82 \times 10^{22}$	$8.64 \pm 2.53 \times 10^{22}$	0.67	1.8×10^2
S_3	$8.02 \pm 1.96 \times 10^{-2}$	$6.4 \pm 3.3 \times 10^{18}$	$5.09 \pm 2.65 \times 10^{19}$	0.38	2.4×10^5
S_4	$3.61 \pm 0.76 \times 10^{1}$	$1.26 \pm 0.56 \times 10^{16}$	$2.57 \pm 1.14 \times 10^{17}$	0.56	6.3×10^7
S_5	$3.13 \pm 0.84 \times 10^{4}$	$1.30 \pm 0.75 \times 10^{13}$	$1.35 \pm 0.78 \times 10^{15}$	0.37	1.7×10^{11}

TABLE S1. Nucleation rates extracted from WTmetaD simulation, *p-value* associated to the Kolmogorov-Smirnov statistic[1], and average acceleration factor.

---

a)Electronic mail: m.salvalaglio@ucl.ac.uk

The characteristic time convergence with the number of independent samples been reported in Fig. S1.
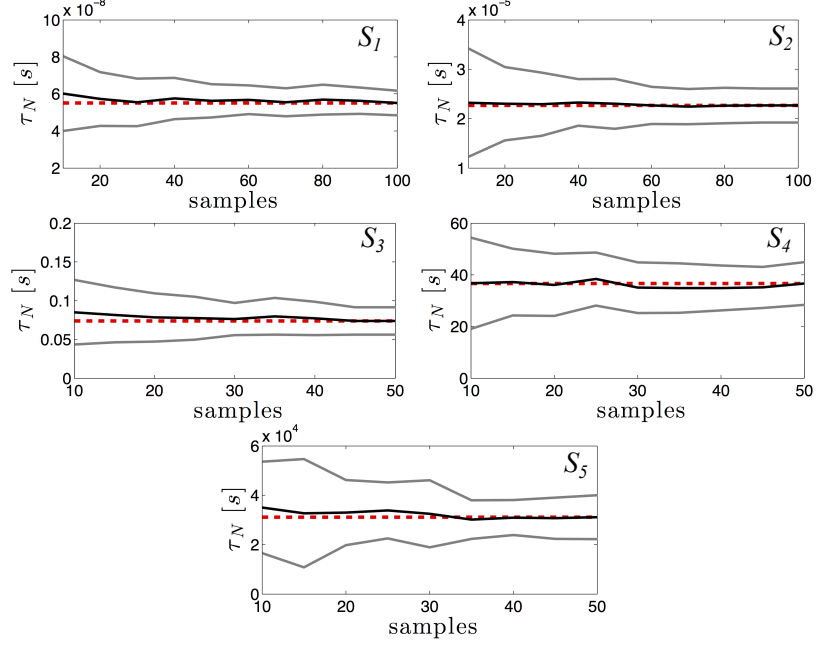

FIG. S1. Convergence of the characteristic time with the number of independent samples.

Errorbars associated to the characteristic time are computed as the standard deviation of $\bar{\tau}$ obtained from a bootstrap-like analysis. The convergence of the errorbars as a function of the number of the bootstrap samples is displayed in Fig. S2.
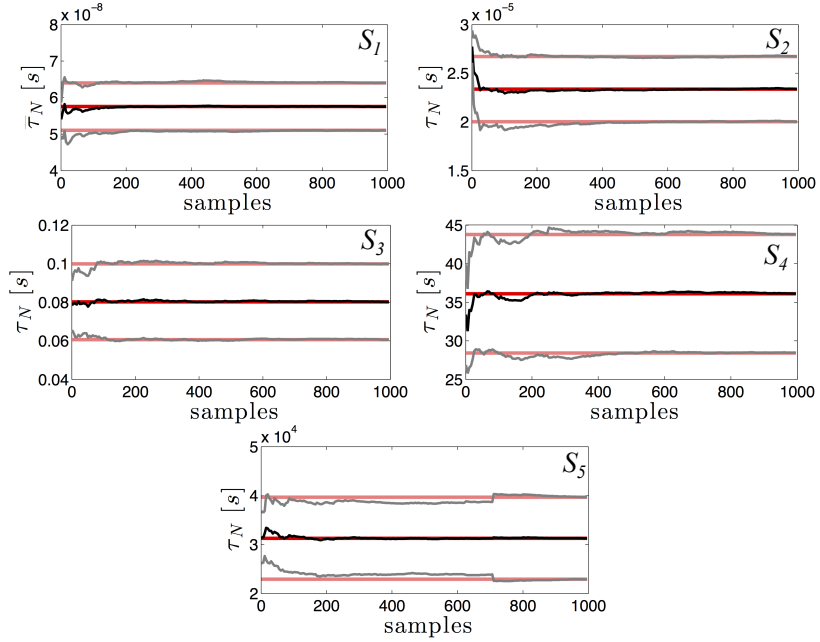

FIG. S2. Convergence of the error estimation as a function of the number of bootstrap samples.

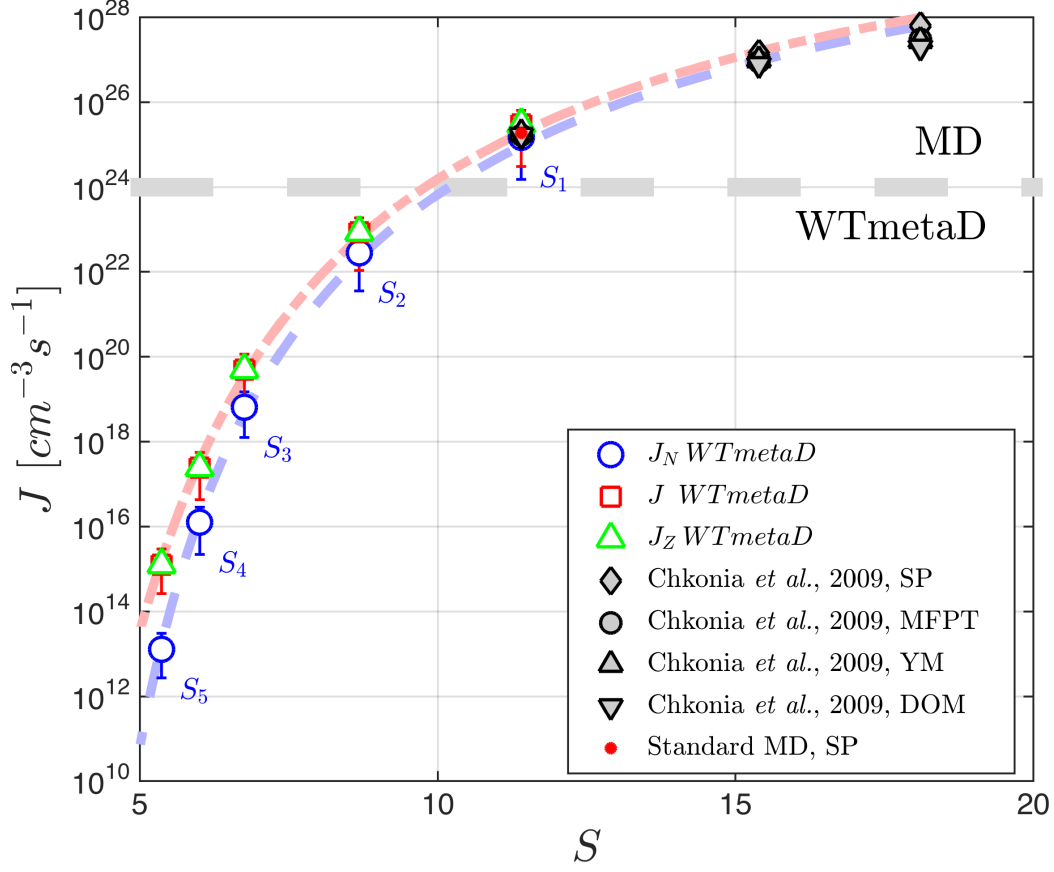## II. IMPORTANCE OF THE ZELDOVICH FACTORS RATIO

FIG. S3. Nucleation rates obtained from WTmetaD simulation (J_N), corrected for finite-size effects using Eq. 20 (J), and corrected for finite size explicitly accounting for the ratio of Zeldovich factors (J_Z). It can be seen that J_Z values are well within the errorbar of the J values. Literature data[2] obtained with different post-processing techniques at high supersaturation has been reported (SP survival probability, MFPT mean first passage time, YM Yasuoka-Matsumoto, DO direct observation) together with an independent calculation of the nucleation rate at S=11.4 performed applying the SP method to standard MD simulations.

III. SURFACE TENSION

Fitting the WTmetaD characteristic times allows to compute surface tension γ. In our case the surface tension is $\gamma = 16.9$ mN/m. Our estimate agrees well with the trend recently reported in the literature[3] for a truncated Lennard-Jones potential. In order to appreciate such agreement, in Fig S4 the σ value computed at 80.7 K is reported together with surface tension values of Ref. 3.
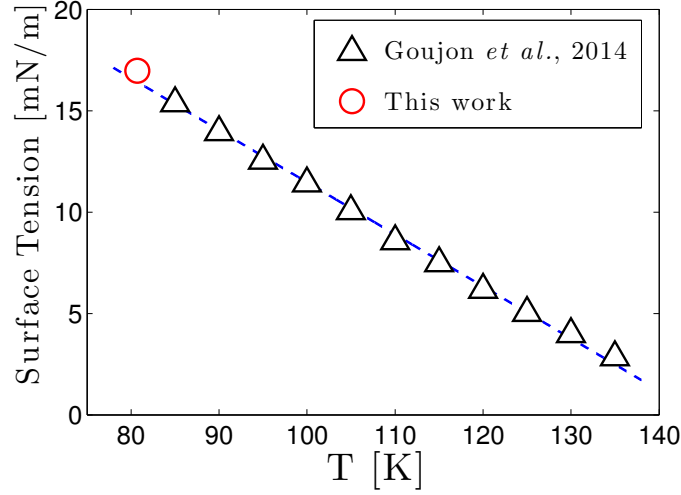

FIG. S4. Surface tension as a function of temperature. The value calculated from the fitting of the nucleation rates J is reported together with the data from Goujon *et al.*[3].

[1]M. Salvalaglio, P. Tiwary, and M. Parrinello, "Assessing the reliability of the dynamics reconstructed from metadynamics," *J. Chem. Theory Comput.*, vol. 10, no. 4, pp. 1420–1425, 2014.

[2]G. Chkonia, J. Wölk, R. Strey, J. Wedekind, and D. Reguera, "Evaluating nucleation rates in direct simulations," *J. Chem. Phys.*, vol. 130, no. 6, pp. –, 2009.

[3]F. Goujon, P. Malfreyt, and D. J. Tildesley, "The gas-liquid surface tension of argon: A reconciliation between experiment and simulation," *J. Chem. Phys.*, vol. 140, no. 24, 2014.